健康传播中的社交机器人：特征、影响与治理

王乙朵　方　伟
（北京信息科技大学公共管理与传媒学院，北京　100192）

摘　要： 近年来社交机器人广泛活跃在网络之中，已经成为网络环境的一部分。2019 年的一项研究报告显示，71% 提到美国股市趋势的推特用户，很可能是机器人 [1]；我国学者张洪忠等发现，在推特有关中美贸易谈判议题的讨论中，社交机器人占 13%，发布内容占比接近 20%。[2] 社交机器人既让网络信息生态开始向“人机共生”转变，也因传播虚假信息、扰乱政治、危害公共健康等负面影响受到了各界的广泛关注。本文主要关注健康传播领域，围绕当前备受关注的社交机器人，在梳理社交机器人的定义、特征和对健康传播所产生的影响的基础上，提出相应的治理策略。

关键词： 健康传播；社交机器人；特征、影响与治理；传播策略与影响　　　**中图分类号：** TN986　　　**文献标识码：** A
文章编号： 1671-0134（2022）06-074-03　　　DOI：10.19483/j.cnki.11-4653/n.2022.06.022
本文著录格式： 王乙朵，方伟 . 健康传播中的社交机器人：特征、影响与治理 [J]. 中国传媒科技，2022（06）：74-76.

1. 社交机器人的内涵与特征

社交机器人是完全或部分由计算机算法控制的社交媒体账户，这些账户通过设置首页照片或资料来模仿人类用户，并通过转发、点赞来模仿人类行为。它们可以自动生成内容，并与人类互动；也可以被用来传播虚假信息，以此改变人们对政治问题的看法。

然而，对社交机器人，目前仍然缺乏一个精确且被广泛认同的定义，来自不同领域的研究人员从多个的角度定义了社交机器人。计算机科学家和工程师倾向于从技术角度定义社交机器人，关注活动水平、自动化程度、算法和人工智能的使用等特性。[3] 社会科学领域的研究者通常对社会影响更感兴趣，比如张洪忠等认为社交机器人是一种拥有身份、人格属性等特征的虚拟人工智能形象，它可以在社交网络中与人互动。[4] 郑晨予等认为社交机器人是由程序员建立并由算法控制社交媒体账号集群。[5]

早期的社交网络开始，社交机器人就与人类共存。早期的社交机器人与垃圾邮件系统类似，它们功能单一，仅支持自动发布内容 [6]，近几年随着大数据、云计算、深度学习等计算机技术的不断发展，社交机器人的功能逐渐增多，影响也逐渐加大，分布也更加广泛。据估计，2017 年推特的活跃账户中，机器人平均占 15%[7]；2019 年脸书的活跃账户中，机器人平均占 11%。[8]

不同的社交机器人具有不同的特征。简单的社交机器人只负责自动发布或转发一种类型的内容，它们不会试图歪曲所发布的内容，也不存在不良的动机。[9] 如推特账户 @big_ben_clock，它会模仿真正的大笨钟每小时发布一条消息。复杂的社交机器人采用各种策略来模拟人类用户，以此获得信任和关注。[10] 这类社交机器人影响着网络生态系统，是当前各界研究和检测的重点。还有一类社交机器人的活动痕迹没有前两者明显，它们是虚假粉丝。这些虚假粉丝通常会互相关注，形成一个社交网络，为每个账户增加可信度，让它们避免因缺乏粉丝而被平台标记。这类账户通常采取真人和自动化操作相结合的方法运行，Cresci 等的一项研究显示，无论是人类还是监督机器学习算法都无法成功识别这类机器人。[11]

目前，国外关于机器人特征、检测和影响的新研究层出不穷。意大利学者 Stefano Cresci 发现，关于社交机器人的研究数量自 2014 年开始激增。[12] 但我国目前的相关研究较少，主要集中于对社交机器人特征、影响和检测方法的研究，或是从中美贸易摩擦等具体事件中展开对社交机器人的实证研究，对社交机器人传播策略的研究相对较少。

2. 健康领域中社交机器人的传播策略与影响

总体来看，社交机器人的传播策略有以下三点。首先，机器人账户活跃在传播的早期阶段，它们会在第一次虚假信息发布后的几秒钟内完成转发，这种传播策略让许多用户接触到虚假信息，有效地促进了病毒式传播。其次，这些机器人账户会回复和提及有影响力的用户，例如一个账户在 18 条推文中提到了 @realDonaldTrump。社交机器人通过这种方式可以让记者、政客等有影响力的人接触到自己的观点，并制造出这种观点被广泛分享的意见环境，以此增加这些名人传播此观点的机会。最后，社交机器人可能会虚报它们的地理位置，同时还会出现在不同的政治派别中，以此获得更广泛的支持。印第安纳大学的一项研究表明，虽然目前没有选民受到这些虚假信息影响的直接证据，但接收社交机器人所传播的虚假信息的州，往往会产生更令人惊讶的选举结果。[13]

涉及健康内容的传播可以统称为健康传播。[14] 在健康传播领域，吸烟、疫苗接种和新冠疫情等议题中均发现了社交机器人存在的痕迹。

第一，通过歪曲在线信息，造成用户认知偏差，以增加有损身体健康产品的销售。社交机器人通过发布大量的、有特定情感倾向的信息，歪曲受众的信息环境，实现对受众认知的影响与行为的引导。比如，2017 年一项利用社交机器人进行电子烟推广的研究发现，社交机器人账户发布电子烟可用于戒烟的内容数量，达到了人类用户发布内容的两倍，而且社交机器人更倾向于用积极的词汇来

描述电子烟，以构建电子烟产品的正面形象。[15]

第二，通过传播谣言，降低人们开展诸如疫苗接种等健康实践的意愿。比如，近年来随着一些国家反疫苗运动的兴起，社交媒体也成为反疫苗人士传播虚假信息或不实信息的阵地。2019年，几位记者进行了一项实验，以“vaccine”为关键词在社交媒体平台脸书（Facebook）进行搜索，结果显示反疫苗言论占主导地位。[16] 其中，社交机器人也已成反疫苗人士传播谣言，混淆舆论环境，实施疫苗接种舆论与行为引导的工具。鉴于当前社交媒体平台是人们获取信息的主要渠道之一，人们在其中所接触的有关疫苗副作用和接种后果的信息，极有可能会影响接种疫苗的决定。[17]

第三，在突发公共卫生事件中散布谣言，造成公众恐慌，威胁社会稳定。以新冠疫情为例，虽然那些转发自主流媒体和可信的信息源的社交机器人，发挥了正面作用[18]，但不可否认的是，由社交机器人充当关键性传播节点所展开的计算宣传已经成为疫情相关谣言盛行的根源[19]，美国卡内基梅隆大学的研究员Young在分析了2亿多条讨论COVID–19的推文后发现，与之前的危机和选举相比，此次疫情中发现了多达两倍的社交机器人活动痕迹，而在他们收集的数据集中，最具影响力的50名转发者中有82%是机器人账户。[20] 而由社交机器人发布的虚假信息，由于较低的成本与较大的发布量，对于加剧谣言传播，制造社会恐慌有着不容忽视的影响。

3. 社交机器人的治理策略

3.1 技术层面提升针对社交机器人的检测能力

第一个针对在线社交网络中自动化账户检测的研究始于2010年1月[21]，当时的机器人检测工作基于监督机器学习和对个人账户的分析，检测时需要对每个账户进行检测，并为其贴上是人类或机器人的二进制标签。[22] 在早期的监督机器学习方法中，每个被调查的账户需要分配一名监督员。这些方法的缺点是，标签只能由操作人员手动分析数据给出，由于对社交机器人的定义不同，不同的操作人员会产生不同的标签方案[23]，而人与机器人之间的灰色地带更是难以定义。

2012年至2013年，几个不同的团队提出了新的检测方案。这些新方案的主要特征是将检测目标作为一个整体，而不是单个账户。这种设计的基本原理是，机器人会与其他机器人协同展开行动，以此形成网络来扩大它们的影响力。[24] 这种用于发现群体协同行为的技术比分析单个账户的效率更高，而通过分析大量的账户，监测系统还可以获取更多的数据，为强大的人工智能算法提供支持。在2018年，新提出的基于群体的检测器数量首次超过了基于个人账户分析的检测器数量。[25]

2013年，对无监督机器学习方法的探索开始出现。无监督学习可以直接对数据建模，这种模式能够直接用于标记样本，不但可以降低人工成本，还能减少人工干预提高样本标记的准确性。[26] 还有一种基于众包的检测方式，它们将检测任务分发给多个标注人员，通过投票决策的方式判断该用户是否为恶意社交机器人。但是这一方法的准确率会随着工作时间的增加而降低，也有一些其他缺陷，如众包工作者的资质和能力审查问题[27]、人工成本问题和用户个人信息保护问题。[28]

在未来，应该在信息传播的早期阶段对社交机器人进行检测，尽早与社交机器人接触可以减轻其影响。此外，随着模型复杂性的不断增加，拥有基准、框架和参考数据集等标准变得越来越重要，因此在评估机器人检测系统时，应该重视技术的通用性和算法偏差。同时，提升检测技术的社会协作水平，将打击社交机器人的行动置于社会协同的框架，在提升研究人员的研究能力，以及相关检测工作人机协作程度的同时，还应重视如何提高网络用户媒介素养这一问题。

3.2 提升学界与业界协同治理水平

基于学术研究开发的检测工具，为社交机器人检测提供理论与技术支持。例如，印第安纳大学开发了社交机器人检测工具Botometer，它基于一种监督机器学习方法[29]，通过对账户打分来区分机器人和人类。在过去的几年中Botometer的应用范围不断扩大，为包括皮尤研究中心在内的众多研究机构提供数据支持。但这些工具的使用也存在一些不确定性，算法有时会将人类用户判定为机器人，进而对其产生禁言、封号等限制措施，这引发了人们对于审查制度合理性的担忧。因此，必须增加有关检测工具准确性的相关研究。

平台应该与学界合作，借助高校科研优势完善自身检测系统，并根据自身技术、用户、数据等结构，对现有检测工具进行改进与提升，开发出适合自身平台的检测工具，并将其融入平台整体产品、运营等工作环节，提升检测的准确性与有效性。此外，平台和学界还可以建立联合辟谣平台，辟谣平台要积极整合政府有关部门、健康专家、主流媒体等机构和社会资源，构筑线上线下多种力量的互动平台，对虚假信息进行协同治理。[30]

3.3 强化对社交机器人的立法监管

在法律规制方面，已有政府出台针对社交机器人规制法案的先例。比如，2019年美国加州政府颁布SB1001法案，要求所有试图影响加州居民投票或购买行为的机器人应该为自己贴上标签，公开自己的机器人身份。但该法案由于对社交机器人的定义模糊、没有让平台方承担披露责任、一些相关资本巨头为自身利益阻碍立法等原因，要有效实施这项监管措施可能还需要一段时间。[31] 因此在管理社交机器人的政策真正有效实施之前，相关部门应该发挥一定的作用，如在健康传播中，公共卫生部门应该在打击虚假健康信息传播和打击不健康产品的网络推广方面发挥作用。[32]

结语

本文梳理了社交机器人的定义、特征、社交机器人在健康传播中的影响及治理策略。如今社交机器人广泛活跃在网络之中，已经成为网络环境的一部分，它们既让网络信息生态开始向“人机共生”转变，也因传播虚假信息、扰乱政治、危害公共健康等负面影响受到了各界的广泛关注，而对这些社交机器人的检测和规制一直是各界关注的研究重点。

在健康传播领域中，吸烟、疫苗接种和新冠疫情等议题中均发现了社交机器人存在的痕迹。值得一提的是，当前的一些研究发现社交机器人在健康传播中并非只会

产生负面影响，比如它们在疫情中发挥着提高官方信息传播效率的作用，这也提醒研究者应该重新审视社交机器人所带来的社会影响。媒

参考文献

[1] Cresci S , Lillo F , Regoli D , et al. Cashtag piggybacking: uncovering spam and bot activity in stock microblogs on Twitter[J]. arXiv e-prints，2018.

[2] 张洪忠，赵蓓，石韦颖：社交机器人在 Twitter 参与中美贸易谈判议题的行为分析 [J]. 新闻界，2020（2）：46-59.

[3] Morstatter，F.，Carley，K.M.，& Liu，H，（2016）.Bot Detection in Social Media：Networks，Behavior，and Evaluation[EB/OL].Available at https：//isi.edu/~fredmors/bottutorial/Tutorial.pdf（2016-04-14）[2018-10-03].

[4] 张洪忠，段泽宁，韩秀：异类还是共生：社交媒体中的社交机器人研究路径探讨 [J]. 新闻界，2019（2）：10-17.

[5] 郑晨予，范红：从社会传染到社会扩散：社交机器人的社会扩散传播机制研究 [J]. 新闻界，2020（3）：51-62.

[6] Tynan，D.Social Spam is Taking Over the Internet[EB/OL].Available at www.itworld.com/article/2832566/it-management/social-spam-is- taking-over-the-internet.html.（2012-04-02）[2018-10-03].

[7] Varol O , Fe Rrara E , Davis C A , et al. Online Human-Bot Interactions： Detection，Estimation，and Characterization[J]. 2017.

[8] Zago M , Nespoli P , Papamartzivanos D , et al. Screening Out Social Bots Interference： Are There Any Silver Bullets?[J]. IEEE Communications Magazine，2019（8）：98-104.

[9] Stefanie，Haustein，Timothy，et al. Tweets as impact indicators： Examining the implications of automated “bot” accounts on Twitter[J]. Journal of the Association for Information Science and Technology，2016.

[10] Yang，Kai - Cheng，Varol O，Davis C A，et al. Arming the public with artificial intelligence to counter social bots[J]. Human Behavior & Emerging Technologies，2019.

[11] Cresci S , Pietro R D , Petrocchi M , et al. Fame for sale： efficient detection of fake Twitter followers[J]. Decision Support Systems，2015，80：56-71.

[12] Cresci S . A Decade of Social Bot Detection[J]. 2020.

[13] Shao C , Ciampaglia G L , Varol O , et al. The spread of fake news by social bots[J]. 2017.

[14] Rogers，Everett M . The field of health communication today.[J]. American Behavioral Scientist，1994.

[15] Allem J P , Ferrara E , Uppu S P , et al. E-Cigarette Surveillance With Social Media Data： Social Bots，Emerging Topics，and Trends[J]. JMIR Public Health and Survllance，2017（4）：98.

[16] Malicious bots and trolls spread vaccine misinformation - now social media companies are fighting back[EB/OL].https://theconversation.com/malicious-bots-and-trolls-spread-vaccine-misinformation-now-social-media-companies-are-fighting-back-123430.

[17] Betsch C , Brewer N T , Brocard P , et al. Opportunities and challenges of Web 2.0 for vaccination decisions[J]. Vaccine，2012（25）：3727-3733.

[18] Al-Rawi A . Bots as Active News Promoters： A Digital Analysis of COVID-19 Tweets[J]. Information（Switzerland），2020（10）.

[19] 史安斌，杨晨晞：信息疫情中的计算宣传：现状、机制与成因 [J]. 青年记者，2021（3）：93-96.

[20] Virginia Alvino Young，（2020） Nearly Half of the Twitter Accounts Discussing ’Reopening America’ May Be Bots[EB/OL].https：//www.scs.cmu.edu/news/nearly-half-twitter-accounts-discussing-reopening-america-may-be-bots，2020（5）.

[21] Yardi S , Romero D M , Schoenebeck G l Detecting spam in a Twitter network[J]. First Monday，2010（1）.

[22] Cresci S . A Decade of Social Bot Detection[J]. 2020

[23] Grimme C , Preuss M , Adam L , et al. Social Bots：Human-Like by Means of Human Control?[J]. Big Data，2017（4）：279.

[24] Zhang J , Zhang R , Zhang Y , et al. The Rise of Social Botnets： Attacks and Countermeasures[J]. IEEE Transactions on Dependable & Secure Computing，2016（99）：1.

[25] Cresci S . A Decade of Social Bot Detection[J]. 2020.

[26] 刘蓉，陈波，于泠，刘亚尚，陈思远 . 恶意社交机器人检测技术研究 [J]. 通信学报，2017（S2）：197-210.

[27] 陈霞，闵华清，宋恒杰 . 众包平台作弊用户自动识别 [J]. 计算机工程，2016（8）：139-145.

[28] Elovici Y , Fire M , Herzberg A , et al. Ethical Considerations when Employing Fake Identities in Online Social Networks for Research[J]. Science and Engineering Ethics，2014，20（4）：1027-1043.

[29] CA Davis，Varol O , Ferrara E , et al. BotOrNot： A System to Evaluate Social Bots[J]. 2016.

[30] 王胜源 . 新媒体背景下伪健康信息的传播与治理——以果壳网“流言百科”证伪的医学健康类信息为例 [J]. 科技传播，2015（22）：110-112+89.

[31] Renee Diresta，A New Law Makes Bots Identify Themselves—That’s the Problem[EB/OL]. https：//www.wired.com/story/law-makes-bots-identify-themselves/，2019（7）.

[32]Jon-PatrickA，EmilioF.CouldSocialBotsPoseaThreattoPublicHealth?[J].Americanjournalofpublichealth，2018（8）：1005.

作者简介：王乙朵（1997-），女，天津，硕士研究生，研究方向：智能舆情；方伟（1981-），女，河北遵化，博士，讲师，研究方向：智能媒体传播、科学传播。

（责任编辑：张晓婧）